

CENSURATION PAR INTERVALLE DES DONNÉES SUR LE RENONCEMENT AU TABAC PROVENANT DE L'ENQUÊTE NATIONALE SUR LA SANTÉ DE LA POPULATION

M. E. Thompson¹ et N. Pantoja Galicia²

RÉSUMÉ

Les données recueillies lors des trois premiers cycles de l'Enquête nationale sur la santé de la population permettent d'appliquer les méthodes d'analyse de données longitudinales à l'étude du commencement de l'usage du tabac, du renoncement au tabac et de la rechute. L'une des questions étudiées est la relation entre le renoncement au tabac et la grossesse. Si une répondante longitudinale qui fumait lors du premier cycle a cessé de fumer lors du deuxième, nous connaissons le moment du renoncement au tabac à un maximum d'un an près, puisqu'on demande à la répondante l'âge auquel elle a cessé de fumer et qu'on connaît sa date de naissance. Nous savons aussi si elle était enceinte au moment du deuxième cycle et si elle a donné naissance depuis l'entrevue du premier cycle. Pour de nombreux sujets, nous pouvons fixer assez exactement la date de conception. Si nous connaissons exactement le moment du renoncement au tabac et la période de la grossesse pour chaque membre du panel longitudinal qui a vécu l'un ou l'autre de ces événements entre deux cycles, nous pourrions modéliser leur lien temporel d'après leurs intensités. Dans la présente communication, nous considérons l'effet de la censure par intervalle des données relatives au moment du renoncement au tabac sur l'estimation des paramètres d'intensité.

MOTS-CLÉS : enquêtes longitudinales, ordonnancement temporel, analyse de données de survie.

1. ENQUÊTES LONGITUDINALES ET SÉQUENCES D'ÉVÉNEMENTS

Dans les études par observation, il est rarement possible d'établir une causalité. Parallèlement, l'interprétation causale d'une association entre des événements peut parfois devenir plus plausible si l'un des événements a tendance à précéder l'autre de près dans le temps. Les enquêtes longitudinales nous permettent d'observer, entièrement ou partiellement, les séquences d'événements concernant des individus.

L'Enquête nationale sur la santé de la population (ENSP) est l'enquête longitudinale nationale de Statistique Canada portant sur la santé et le comportement influant sur la santé. À ce jour, on a recueilli quatre cycles de réponses de l'échantillon longitudinal. Les trois premiers, recueillis en 1994-1995, 1996-1997 et 1998-1999, sont disponibles auprès des Centres de données de recherche. Le plan d'échantillonnage du premier cycle consistait à choisir au hasard un individu par ménage dans environ 17 000 ménages canadiens, et 14 619 sujets ont répondu lors des trois premiers cycles. Ces individus constituent l'échantillon longitudinal jusqu'en 1999. Nous disposons donc d'un grand nombre de sujets pour lesquels il est possible d'observer les séquences d'événements d'après les réponses qu'ils fournissent à des intervalles de deux ans.

En principe, les cycles de l'Enquête nationale sur la santé de la population permettent d'appliquer les méthodes d'analyse de données longitudinales à des événements ponctuels tels que le commencement de l'usage du tabac, le renoncement au tabac et la rechute. Parallèlement, ces événements ont tendance à être censurés par intervalle. Par exemple, l'une des questions étudiées est la relation entre l'usage du tabac et la grossesse. Si une répondante longitudinale qui fumait lors du n^{e} cycle a cessé de fumer lors du $(n+1)^{\text{e}}$ cycle, nous connaissons le moment du renoncement au tabac (en supposant une réponse exacte) à un maximum d'un an près, puisqu'on demande à la répondante l'âge auquel elle a cessé de fumer et qu'on connaît sa date de naissance. Nous savons aussi si elle a déclaré être enceinte au moment du $(n+1)^{\text{e}}$ cycle et si elle a donné naissance depuis l'entrevue du n^{e} cycle. Pour de nombreux sujets, nous pouvons relier le dossier à un dossier des ménages montrant les dates de naissance, et nous pouvons inférer un intervalle relativement faible pour la date de conception.

¹ Université de Waterloo, Waterloo (Ontario), Canada N2L 3G1 (methomps@uwaterloo.ca)

² Université de Waterloo, Waterloo (Ontario), Canada N2L 3G1 (npantoja@uwaterloo.ca)

La présente communication comporte deux objectifs. Premièrement, nous aborderons quelques notions d'association et d'ordonnement temporels et le concept selon lequel un type d'événement en « déclenche » un autre. Deuxièmement, nous décrirons la construction de tests portant sur ces relations temporelles et montrerons comment on peut appliquer les tests au contexte du renoncement au tabac dans le cadre de l'ENSP.

2. ASSOCIATION ET MODÉLISATION CAUSALE

D'après un résumé des écrits de Bradford Hill (1937, 1965) rédigé par Cox (1992), il existe des circonstances dans lesquelles un « effet » obtenu lors d'une étude par observation pourrait faire l'objet d'une interprétation causale : l'effet

- (a) est important,
- (b) est reproduit lors d'études indépendantes,
- (c) présente une relation monotone avec la « dose »,
- (d) correspond à une « expérience naturelle »,
- (e) se comporte logiquement lorsque la cause potentielle est appliquée, supprimée puis réappliquée,
- (f) est cohérent avec ce qu'on sait du sujet à l'étude ou
- (g) est prédit par une théorie raisonnablement bien établie.

Il n'est pas mentionné que l'effet est cohérent avec l'attribution du sujet, probablement parce que Bradford Hill ne pensait pas aux causes du comportement. Dans l'exemple de l'ENSP, nous avons une certaine attribution du sujet, et nous devons décider d'en tenir compte ou non.

L'ordre constitue un facteur important de l'application d'une interprétation causale à une association. Cox (1992) évoque les événements associés C et E où C est une cause candidate de E :

[Pour inférer qu'il est « fallacieux » de considérer C comme une cause (candidate) de E parce que E est indépendant de C , étant donné B] il faut restreindre les variables C , E et B , ne serait-ce que pour exprimer l'asymétrie essentielle entre la cause et l'effet. Le plus simple est d'affirmer que dans le temps, B survient avant C , lequel survient à son tour avant E . (...) On peut parfois ordonner les effets selon la proximité spatiale, plutôt que selon l'ordonnement temporel. Une troisième possibilité consiste à utiliser la connaissance du sujet à l'étude pour présumer un ordonnancement causal des variables.

Il est naturel de se demander si les données longitudinales, dont l'ordre temporel va de soi, sont vraiment essentielles à l'inférence. Lorsqu'on cherche à distinguer les relations entre les variables, est-il parfois utile de pouvoir vérifier la *tendance* de E à suivre C et la *tendance* de C à suivre B ?

Pour démontrer que l'*ordonnement* est un type d'*association*, nous proposons dans la section suivante quelques caractérisations de chacun à l'égard de paires de durées T_1 et T_2 , ainsi que de séquences d'événements de deux types.

3. ASSOCIATION ET ORDONNEMENT TEMPORELS

3.1 Distributions conjointes de durées

Nous pourrions dire que les durées T_1 et T_2 sont *faiblement associées* si elles sont positivement corrélées.

Des conditions suffisantes pour qu'existe une faible association seraient que $T_1 = Z + \varepsilon_1$, $T_2 = Z + \varepsilon_2$, où, Z ,

ε_1 et ε_2 sont positifs et indépendants ou, de manière plus générale, que la distribution conjointe de T_1 et T_2 soit un mélange convenable de distributions conjointement indépendantes. Dans le premier cas, surtout, il est naturel de considérer T_1 et T_2 comme associées, car toutes deux sont précédées (strictement) par une autre durée Z , candidate d'un événement causal. Dans le deuxième cas, on peut être en présence d'un mécanisme semblable où l'ordonnement n'est pas strict.

Nous pourrions dire que les durées T_1 et T_2 sont *fortement associées* si nous savions qu'une borne inférieure de T_1 déplace davantage la masse de T_2 , c.-à-d. si, pour chaque valeur de $t_1, t_2 > 0$, les fonctions conditionnelles de survie satisfont

$$\Pr(T_2 > t_2 | T_1 > t_1) > \Pr(T_2 > t_2) \quad (1)$$

On constate aisément que cette équation équivaut à

$$F(t_1, t_2) > F_1(t_1)F_2(t_2) \quad (2)$$

pour toutes les valeurs de $t_1, t_2 > 0$, dont on peut montrer qu'elles supposent une faible association.

Une condition suffisante pour qu'existe une forte association serait que T_1 soit exponentielle et que $F_2(t_2 | T_1 = t_1)$ augmente avec t_1 .

L'ordonnement temporel de distributions conjointes de durées est, en quelque sorte, une association, en l'occurrence une association « unilatérale ».

La première définition de l'ordonnement est incontestablement celle d'une relation d'ordre. Nous disons que T_1 est un *précurseur strict* de T_2 si $T_2 - T_1$ est une variable aléatoire non négative.

Si T_1 est un précurseur strict de T_2 , T_1 et T_2 sont faiblement associées lorsque $T_1, T_2 - T_1$ sont indépendantes, et fortement associées lorsque T_1 est aussi exponentielle. Comme on rencontre rarement un ordonnement strict dans le genre d'application qui nous occupe, il convient d'envisager d'autres définitions.

Les tests non paramétriques peuvent reposer sur la notion de « précurseur faible ».

T_1 est un *précurseur faible* de T_2 si $P(T_1 < T_2)$ selon la distribution conjointe est supérieure à $P(T_1 < T_2)$ selon l'hypothèse de l'indépendance de T_1 et T_2 .

À l'instar de la forte association liée aux distributions de durées, il existe une relation de « précurseur fort ».

T_1 est un *précurseur fort* de T_2 si

$$F_2(t_1 | T_1 = t_1) > F_2(t_1) \quad (3)$$

pour toutes les valeurs de t_1 .

Il n'est pas difficile de montrer qu'un précurseur au sens fort est aussi un précurseur au sens faible. Parallèlement, il pourrait ne pas sembler assez fort. Plus T_2 a tendance à suivre de près T_1 , plus un lien causal semblerait plausible. Il existe également un concept de « précurseur rapproché » qui ne suppose aucun autre et qui n'est supposé par aucun autre, mais qui suppose une association locale de T_1 et T_2 .

T_1 est un *précurseur rapproché* de T_2 si, pour les nombres positifs δ et $\kappa(t_1)$, nous avons

$$\frac{F_2(t_1 + \kappa(t_1) | T_1 = t_1)}{F_2(t_1 | T_1 = t_1)} < \frac{F_2(t_1 + \kappa(t_1))}{F(t_1)} - \delta \quad (4)$$

3.2 Intensités conjointes

L'association temporelle d'événements récurrents sans liens logiques forts est très naturellement une propriété *locale*; il en existe une version conditionnelle et une version « marginale ».

Version marginale :

Prenons un processus ponctuel comportant deux sortes d'événements, \times et o . Supposons que les deux sortes d'événements présentent une fonction conjointe d'intensité inconditionnelle $\lambda(t_1, t_2)$.

Nous pourrions dire que les événements \times et o sont *associés dans le temps* si, pour chaque valeur de t_1 et pour chaque valeur de t_2 comprise à l'intérieur d'une certaine distance $\kappa(t_1)$ de t_1 , nous avons

$$\lambda(t_1, t_2) > \lambda_1(t_1)\lambda_2(t_2). \quad (5)$$

L'ordre temporel d'événements récurrents doit supposer la proximité en plus de l'ordre. Dans le cas d'une version « conditionnelle », nous pourrions dire que l'événement o est un *précurseur rapproché* de l'événement \times si, pour des valeurs convenablement choisies de $\kappa(s)$ et $\delta(s) > 0$, nous avons

$$\lambda_2(u | H(s-) \text{ et } o \text{ au temps } s) > \lambda_2(u | H(s-)) + \delta(s) \quad (6)$$

pour u entre s et $s + \kappa(s)$. Ici, $H(s-)$ représente les antécédents du processus conjoint avant le moment s .

Il est possible, mais pas nécessaire, que deux événements o et \times soient des précurseurs l'un de l'autre en ce sens.

4. MODÈLES DÉCLENCHEURS

Le déclenchement est un cas particulier de l'ordonnancement temporel, où un modèle causal est explicite. Il semble tout naturel de présenter une formulation conditionnelle du déclenchement : la voici.

4.1 Distribution conjointe de durées

Nous pourrions dire que T_1 déclenche T_2 si la fin de la durée T_1 modifie la fonction de risque de T_2 . À titre d'exemple simple, supposons que si T_1 représentait l'infini, T_2 présenterait la fonction de survie $F_{02}(t_1)$ et la fonction de risque $\lambda_{02}(u)$. Toutefois, si $T_1 = t_1$, alors T_2 présente la fonction de risque $\lambda_{02}(u)$ avant t_1 et $e^\beta \lambda_{02}(u)$ après t_1 , où $\beta > 0$. Il est possible de montrer que, de manière inconditionnelle, T_2 présente alors la fonction de survie

$$F_2(t) = F_{02}(t) \left[\int_0^t f_1(t_1) e(\beta, t_1, t) dt_1 + F_1(t) \right], \quad (7)$$

où

$$e(\beta, t_1, t) = \exp \left\{ -(e^\beta - 1) \int_{t_1}^t \lambda_{02}(u) du \right\}. \quad (8)$$

À partir de cette expression, il est facile de montrer que T_1 est un précurseur fort de T_2 . À l'égard des durées, il est possible de formuler ce genre de propriété de déclenchement à long terme. Il est aussi possible de formuler le déclenchement local, à titre de cas particulier de la même notion à l'égard des intensités conjointes ci-dessous.

En fonction d'une modification d'échelle à court terme du risque de T_2 , la formulation « locale » donne une relation de densité

$$f_2(t|T_1=t_1) = e^\beta f_{02}(t) e(\beta, t_1, t) \quad (9)$$

pour chaque valeur de t_1 et de t comprise entre t_1 et $t_1 + \kappa(t_1)$. On constate aisément que dans ce cas, T_1 est un précurseur rapproché de T_2 , comme l'exprime l'équation (4).

4.2 Intensités conjointes

Dans un sens local, nous pourrions dire que l'événement \circ déclenche l'événement \times s'il augmente (pendant un certain temps) l'intensité conditionnelle de \circ , étant donné les antécédents récents. Par exemple, supposons que $H(s-)$ représente encore les antécédents du processus conjoint avant le moment s . Alors, \circ déclenche \times si, pour certaines valeurs de $\kappa(s)$, nous avons

$$\lambda_2(u|H(s-) \text{ et } \circ \text{ au temps } s) = e^\beta \lambda_2(u|H(s-)) \quad (10)$$

pour $s < u < s + \kappa(s)$ ($\beta > 0$).

Il est possible que les deux sortes d'événements récurrents soient des déclencheurs l'un de l'autre.

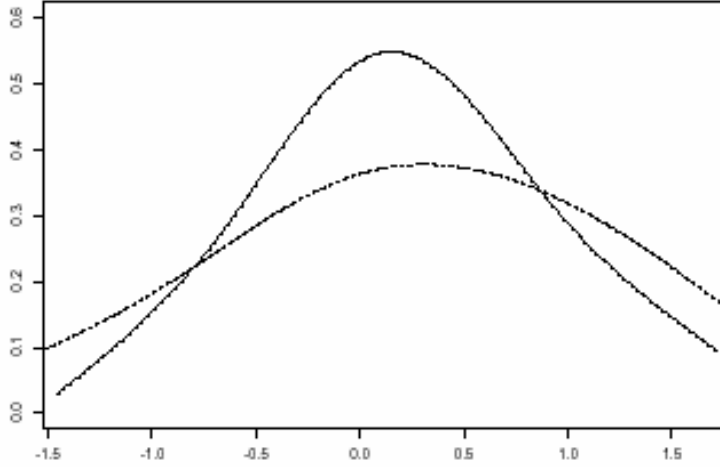
5. TEST NON PARAMÉTRIQUE DE L'ORDRE AVEC DONNÉES COMPLÈTES

Pour effectuer un test non paramétrique simple de l'ordonnement avec des données complètes, il semble nécessaire d'examiner la distribution empirique de $T_2 - T_1$. Rappelons que T_1 est un précurseur faible de T_2 si $P(T_2 - T_1 > 0)$ est plus grand qu'il ne le serait selon l'hypothèse de l'indépendance de T_1, T_2 ; T_1 est un précurseur rapproché de T_2 si la fonction de densité de $T_2 - T_1$, étant donné $T_1 = t_1$ et $T_2 > T_1$, est plus élevée à l'égard d'arguments légèrement positifs que la densité de $T_2 - t_1$, étant donné $T_2 > t_1$. Nous illustrons l'utilisation de $T_2 - T_1$ par un exemple quelque peu artificiel, inspiré par le contexte du renoncement au tabac et de la grossesse.

Supposons que T_1 (correspondant à la conception) présente un taux de risque constant de 0,1. Supposons que T_2 (correspondant au renoncement au tabac) présente un taux de risque de 0,05 avant T_1 et de 0,15 après T_1 pour 0,5 unité de temps et qui revient à 0,05 par la suite. Prenons l'origine temporelle 0 comme moment du n^{e} cycle, et simulons les antécédents des sujets qui, au point d'origine, n'ont connu ni l'événement T_1 ni l'événement T_2 (les fumeuses non enceintes). Supposons que nous voulions modéliser les événements survenant entre le n^{e} et le $(n+1)^{\text{e}}$ cycle.

Lors d'une simulation portant sur 10 000 sujets, 145 présentaient les valeurs $0 < T_1, T_2 < 2$.

Pour vérifier graphiquement si T_1 est un précurseur (rapproché) de T_2 , nous pourrions prendre la distribution (ou la densité) de $T_2 - T_1$, sachant que $0 < T_1, T_2 < 2$. Nous pourrions comparer la densité empirique ou estimée de $T_2 - T_1$, sachant que $0 < T_1, T_2 < 2$, à la densité conditionnelle supposée de $T_2 - T_1$, d'après les marginales estimées de T_1, T_2 et selon l'hypothèse de l'indépendance.



La figure 1 montre la comparaison pour l'ensemble de données simulé. La densité estimée de $T_2 - T_1$ montre une concentration de masse vis-à-vis 0.

Un test non paramétrique formel de l'association, qui remettrait en question l'hypothèse selon laquelle T_1 est un précurseur rapproché de T_2 , pourrait se dérouler comme suit : d'après les données, estimons $F_2(\cdot)$ et $F_1(\cdot)$, ainsi que la distribution conjointe de $(T_1, T_2 - T_1)$. Pour estimer cette dernière, il suffit d'estimer

$$\frac{F_2(t_1 + \kappa(t_1) | T_1 = t_1)}{F_2(t_1 | T_1 = t_1)} \quad (11)$$

pour chaque valeur de t_1 et chaque valeur convenable de $\kappa(t_1)$. Calculons la variable à tester

$$\int (\hat{R}(t_1) - \frac{\hat{F}_2(t_1 + \kappa(t_1))}{\hat{F}_2(t_1)}) d\hat{F}_1(t_1), \quad (12)$$

et comparons-la à deux fois son erreur-type estimée, pour vérifier l'hypothèse nulle selon laquelle sa moyenne est 0.

En principe, il n'est pas difficile d'estimer $F_2(\cdot)$, $F_1(\cdot)$ et la distribution conjointe de $(T_1, T_2 - T_1)$, à l'aide des méthodes d'estimation de la distribution et de la densité pour des enquêtes complexes. Voir, par exemple, Bellhouse et Stafford (1999), Lohr et Buskirk (1999). On pourrait utiliser des méthodes de rééchantillonnage pour évaluer l'erreur-type.

Remarque : L'exemple est artificiel, car l'hypothèse d'un risque constant de renoncement au tabac en l'absence d'une grossesse constituerait une simplification excessive. En fait, lorsque nous observons $T_2 < 2$, nous observons le moment du commencement d'une *tentative de renoncement* au tabac ayant duré assez longtemps pour être déclarée au moment 2. Strictement parlant, le modèle devrait reconnaître la possibilité que le renoncement au tabac soit suivi d'une rechute avant la fin de la période; si cette possibilité n'est pas comprise explicitement, nous devrions au moins laisser le risque du renoncement au tabac déclaré augmenter au-delà de $[0, 2]$.

6. CENSURATION PAR INTERVALLE

Dans les enquêtes longitudinales menées à des moments très espacés, il est possible que le point limite de T_1 ou de T_2 , ou le moment auquel survient \circ ou \times , soit *censuré par intervalle* ou qu'on observe qu'il est compris à l'intérieur d'un intervalle seulement. C'est le cas du moment du renoncement au tabac dans le cadre de l'ENSP. L'intervieweur demande :

Avez-vous déjà fumé des cigarettes à un moment quelconque? (...) À quel âge avez-vous cessé de fumer (des cigarettes) tous les jours? (...) Comparativement à notre interview en [MOIS/ANNÉE], vous dites que vous ne fumez plus. Pourquoi avez-vous cessé?

Si $[0, 2]$ est la période comprise entre les interviews et qu'on observe le renoncement au tabac, on observe donc le moment T_2 du renoncement au tabac entre deux points limites T_{20} et T_{21} , où T_{20} est soit 0 ou un anniversaire de naissance et T_{21} , soit un anniversaire de naissance ou 2. Si les sujets sont des femmes de 15 à 45 ans qui fument régulièrement et qui ne sont pas enceintes au moment 0, nous pouvons supposer que T_1 est le moment de la conception, qui serait idéalement « observé » s'il survenait avant 2.

Si nous avons un modèle déclencheur semi-paramétrique dans le sens de l'équation (9), nous pouvons composer avec cette situation en calculant simplement la fonction de vraisemblance pertinente, que nous utiliserons pour estimer β .

Par exemple, supposons que $\lambda_1(t)$ et $\lambda_{02}(\tau)$ soient respectivement les fonctions de risque pour T_1 et pour T_2 lorsque T_1 est ∞ . Supposons que $\lambda_{12}(u|t; \beta)$ est la fonction de risque pour T_2 à u , étant donné $T_1 = t$, de sorte que $\lambda_{12}(u|t; \beta)$ pourrait être $e^\beta \lambda_{02}(u)$ pour u entre t et $t+0,5$, et $\lambda_{02}(u)$ pour les autres valeurs de u .

En remplaçant 2 par un point limite général a_1 et en supposant que t_{20i} soit a_1 si T_2 est inobservé pour le sujet i , nous pouvons formuler la fonction de vraisemblance $\prod_i L(t_{20i}, t_{21i}, t_{1i})$

où $L(t_{20}, t_{21}, t_1) = :$

1. $A(a_1)$ si $a_1 \leq t_1, t_{20}$
2. $A(t_1)\lambda_1(t_1)dt_1 B_{12}(a_1 | t_1; \beta)$ si $0 < t_1 < a_1 = t_{20}$
3. $\int_{t_{20}}^{t_{21}} A(\tau)\lambda_{02}(\tau) \exp\left\{-\int_{\tau}^{a_1} \lambda_1(u)du\right\} \lambda_1(\tau) d\tau$ si $0 \leq t_{20} < t_{21} \leq a_1 < t_1$
4. $\int_{t_{20}}^{t_1} A(\tau)\lambda_{02}(\tau) \exp\left\{-\int_{\tau}^{t_1} \lambda_1(u)du\right\} \lambda_1(\tau) d\tau dt_1 + A(t_1)\lambda_1(t_1)dt_1 [1 - B_{12}(t_{21} | t_1; \beta)]$ si $0 \leq t_{20} < t_1 < t_{21} \leq a_1$
5. $A(t_1)\lambda_1(t_1)dt_1 [B_{12}(t_{20} | t_1; \beta) - B_{12}(t_{21} | t_1; \beta)]$ si $0 < t_1 < t_{20} < t_{21} \leq a_1$
6. $\int_{t_{20}}^{t_{21}} A(\tau)\lambda_{02}(\tau) \exp\left\{-\int_{\tau}^{t_1} \lambda_1(u)du\right\} \lambda_1(\tau) d\tau dt_1$ si $0 < t_{20} < t_{21} \leq t_1 < a_1$

et

$$A(t) = \exp\left\{-\int_0^t [\lambda_1(u) + \lambda_{02}(u)]du\right\}$$

$$B_{12}(\tau | t; \beta) = \exp\left\{-\int_{\tau}^t \lambda_{12}(u | t; \beta)du\right\}.$$

Des modèles plus réalistes laisseraient le risque λ_1 dépendre de variables aléatoires telles que l'état civil et l'âge, et les risques λ_{02} et λ_{12} dépendre de variables aléatoires telles que le niveau d'instruction et une certaine mesure de la force de la dépendance.

Pour des tests non paramétriques dans l'esprit de l'équation (12), il nous faut adapter l'estimation des fonctions de survie et la distribution conjointe de $T_1, T_2 - T_1$, étant donné $0 < T_1, T_2 < a_1$, en fonction du cas où T_2 est censuré par intervalle. Une approche directe consisterait à construire une estimation de la densité à deux variables pour T_1, T_2 sur le rectangle $(0 < T_1, T_2 < a_1)$, dans le sens de la méthode proposée par Duchesne et Stafford (2002) et à utiliser des noyaux orientés parallèlement à la ligne $t_1 = t_2$, puis à effectuer une intégration pour produire le numérateur et le dénominateur de $\hat{R}(t_1)$ pour chaque valeur de t_1 , ainsi que les autres quantités comprises dans l'équation (12).

REMERCIEMENTS

Cette recherche a bénéficié d'une subvention de recherche du CRSNG. Nous tenons à remercier J. F. Lawless pour ses observations.

RÉFÉRENCES

- Bellhouse, D. et Stafford, J. (1999), "Density Estimation from Complex Surveys", *Statistica Sinica*, 9, pp. 407-424.
- Bradford Hill, A. (1937), *Principles of Medical Statistics*, Arnold.
- Bradford Hill, A. (1965), "The Environment and Disease: Association or Causation", *Proceedings of the Royal Society of Medicine*, 58, pp. 295-300.
- Cox, D.R. (1992), "Causality: Some Statistical Aspects", *Journal of the Royal Statistical Society, Series A*, 155, pp. 291-301.
- Duchesne, R. et Stafford, J.E. (2002), "A Kernel Density Estimate for Internal Censored Data".
- Lohr, S. et Buskirk, T. (1999), "Density Estimation with Complex Survey Data", *Proceedings of the Survey Methods Section*, Statistical Society of Canada, pp. 27-32.